Capacity of a Class of Deterministic Relay Channels

Thomas M. Cover and Young-Han Kim*

Abstract

The capacity of a class of deterministic relay channels with the transmitter input X, the receiver output Y, the relay output $Y_1 = f(X, Y)$, and a separate communication link from the relay to the receiver with capacity R_0, is shown to be

$$C(R_0) = \max_{p(x)} \min\{I(X;Y) + R_0,\ I(X;Y,Y_1)\}.$$

Thus every bit from the relay is worth exactly one bit to the receiver. Two alternative coding schemes are presented that achieve this capacity. The first scheme, "hash-and-forward", is based on a simple yet novel use of random binning on the space of relay outputs, while the second scheme uses the usual "compress-and-forward". In fact, these two schemes can be combined together to give a class of optimal coding schemes. As a corollary, this relay capacity result confirms a conjecture by Ahlswede and Han on the capacity of a channel with rate-limited state information at the decoder in the special case when the channel state is recoverable from the channel input and the output.

1 Introduction with Gaussian Relay

Consider the Gaussian relay problem shown in Figure 1. Suppose the receiver Y and the

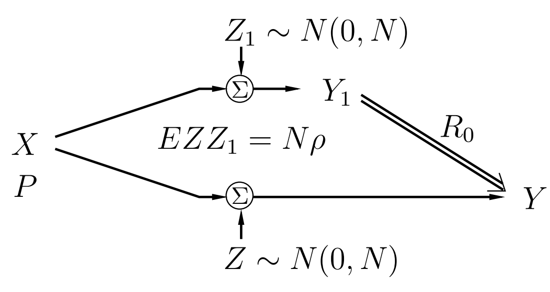

Figure 1: Gaussian relay channel with a noiseless link.

*Email: cover@stanford.edu, yhk@ucsd.edu

relay Y_1 each receive information about the transmitted signal X of power P. Specifically, let

$$Y = X + Z$$
$$Y_1 = X + Z_1,$$

where (Z, Z_1) have correlation coefficient ρ and are jointly Gaussian with zero mean and equal variance $EZ^2 = EZ_1^2 = N$. What should the relay Y_1 say to the ultimate receiver Y? If the relay sends information at rate R_0, what is the capacity $C(R_0)$ of the resulting relay channel?

We first note that the capacity from X to Y, ignoring the relay, is

$$C(0) = \frac{1}{2}\log\left(1 + \frac{P}{N}\right) \quad \text{bits per transmission.}$$

The channel from the relay Y_1 to the ultimate receiver Y has capacity R_0. This relay information is sent on a side channel that does not affect the distribution of Y, and the information becomes freely available to Y as long as it doesn't exceed rate R_0. We focus on three cases for the noise correlation ρ: $\rho = 1, 0$, and -1.

If $\rho = 1$, then $Y_1 = Y$, the relay is useless, and the capacity of the relay channel is $C(R_0) = (1/2)\log(1 + P/N) = C(0)$ for all $R_0 \geq 0$.

Now consider $\rho = 0$, i.e., the noises Z and Z_1 are independent. Then the relay Y_1 has no more information about X than does Y, but the relay furnishes an independent look at X. What should the relay say to Y? This capacity $C(R_0)$, mentioned in [4], remains unsolved and typifies the primary open problem of the relay channel. As a partial converse, Zhang [12] obtained the strict inequality $C(R_0) < C(0) + R_0$ for all $R_0 > 0$.

How about the case $\rho = -1$? This is the problem that we solve and generalize in this note. Here the relay, while having no more information than the receiver Y, has much to say, since knowledge of Y and Y_1 allows the perfect determination of X. However, the relay is limited to communication at rate R_0. Thus, by a simple cut-set argument, the total received information is limited to $C(0)+R_0$ bits per transmission. We argue that this rate can actually be achieved. Since it is obviously the best possible rate, the capacity for $\rho = -1$ is given as

$$C(R_0) = C(0) + R_0.$$

(See Figure 2.) Every bit sent by the relay counts as one bit of information, despite the fact that the relay doesn't know what it is doing.

We present two distinct methods of achieving the capacity. Our first coding scheme consists

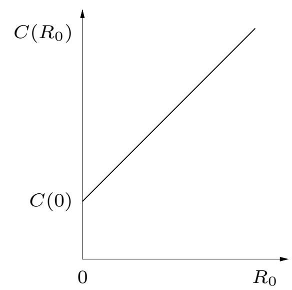

Figure 2: Gaussian relay capacity $C(R_0)$ vs. the relay information rate R_0.

of hashing Y_1^n into nR_0 bits, then checking the $2^{nC(R_0)}$ codewords $X^n(W)$, $W \in 2^{nC(R_0)}$, one by one, with respect to the ultimate receiver's output Y^n and the hash check of Y_1^n. More specifically, we check whether the corresponding estimated noise $\hat{Z}^n = Y^n - X^n(W)$ is typical, and then check whether the resulting $Y_1^n(W) = X^n(W) + \hat{Z}^n$ satisfies the hash of the observed Y_1^n. Since the typicality check reduces the uncertainty in $X^n(W)$ by a factor of $2^{nC(0)}$ while the hash check reduces the uncertainty by a factor of 2^{nR_0}, we can achieve the capacity $C(R_0) = C(0) + R_0$.

It turns out hashing is not the unique way of achieving $C(R_0) = C(0) + R_0$. We can compress Y_1^n into $\hat{Y}_1^n$ using nR_0 bits with Y^n as side information in the same manner as in Wyner–Ziv source coding [11], which requires

$$R_0 = I(Y_1; \hat{Y}_1 | Y).$$

Thus, nR_0 bits are sufficient to reveal $\hat{Y}_1^n$ to the ultimate receiver Y^n. Then, based upon the observation $(Y^n, \hat{Y}_1^n)$, the decoder can distinguish 2^{nR} messages if

$$R < R^* := I(X; Y, \hat{Y}_1).$$

For this scheme, we now choose the appropriate distribution of $\hat{Y}_1$ given Y_1. Letting

$$\hat{Y}_1 = Y_1 + U,$$

where $U \sim N(0, \sigma^2)$ is independent of (X, Z, Z_1), we can obtain the following parametric

expression of $R^*(R_0)$ over all $\sigma^2 > 0$:

$$R^*(\sigma^2) = I(X; Y, \hat{Y}_1) = \frac{1}{2}\log\left(\frac{(P+N)\sigma^2 + 4PN}{N\sigma^2}\right) \tag{1}$$

$$R_0(\sigma^2) = I(Y_1; \hat{Y}_1|Y) = \frac{1}{2}\log\left(\frac{(P+N)\sigma^2 + 4PN}{(P+N)\sigma^2}\right). \tag{2}$$

Setting $R_0(\sigma_0^2) = R_0$ in (2), solving for σ_0^2, and inserting it in (1), we find the achievable rate is given by

$$R^*(\sigma_0^2) = R_0 + \frac{1}{2}\log\left(1 + \frac{P}{N}\right) = C(0) + R_0,$$

so "compress-and-forward" also achieves the capacity.

Inspecting what it is about this problem that allows this solution, we see that the critical ingredient is that the relay output $Y_1 = f(X, Y)$ is a deterministic function of the input X and the receiver output Y. This leads to the more general result stated in Theorem 1 in the next section.

2 Main Result

We consider the following relay channel with a noiseless link as depicted in Figure 3. We

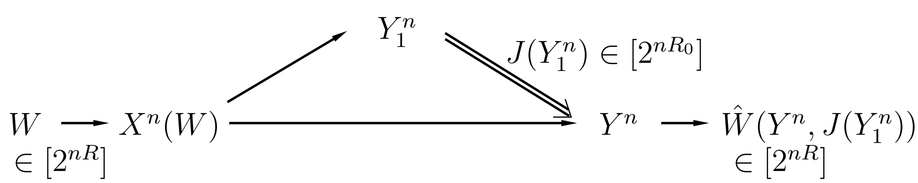

Figure 3: Relay channel with a noiseless link.

define a *relay channel with a noiseless link* $(\mathcal{X}, p(y, y_1|x), \mathcal{Y}\times\mathcal{Y}_1, R_0)$ as the channel where the input signal X is received by the relay Y_1 and the receiver Y through a channel $p(y, y_1|x)$, and the relay can communicate to the receiver over a separate noiseless link of rate R_0. We wish to communicate a message index $W \in [2^{nR}] = \{1, 2, \ldots, 2^{nR}\}$ reliably over this relay channel with a noiseless link.[1] We specify a $(2^{nR}, n)$ code with an encoding function $X^n : [2^{nR}] \to \mathcal{X}^n$, a relay function $J : \mathcal{Y}_1^n \to [2^{nR_0}]$, and the decoding function $\hat{W} : \mathcal{Y}^n \times [2^{nR_0}] \to [2^{nR}]$. The probability of error is defined by $P_e^{(n)} = \Pr\{W \neq \hat{W}(Y^n, J(Y_1^n)\}$, with the message W distributed uniformly over $[2^{nR}]$. The capacity $C(R_0)$ is the supremum of the rates R for which $P_e^{(n)}$ can be made to tend to zero as $n \to \infty$.

[1] Henceforth, the notation $i \in [2^{nR}]$ is interpreted to mean $i \in \{1, 2, \ldots, 2^{nR}\}$.

We state our main result.

Theorem 1. *For the relay channel* $(\mathcal{X}, p(y, y_1|x), \mathcal{Y} \times \mathcal{Y}_1)$ *with a noiseless link of rate* R_0 *from the relay to the receiver, if the relay output* $Y_1 = f(X, Y)$ *is a deterministic function of the input* X *and the receiver output* Y*, then the capacity is given by*

$$C(R_0) = \max_{p(x)} \min\{I(X;Y) + R_0,\ I(X;Y_1,Y)\}.$$

The converse is immediate from the simple application of the max-flow min-cut theorem on information flow [6, Section 15.10].

The achievability has several interesting features. First, as we will show in the next section, a novel application of random binning achieves the cut-set bound. In this coding scheme, the relay simply sends the hash index of its received output Y_1^n.

What is perhaps more interesting is that the same capacity can be achieved also via the well-known "compress-and-forward" coding scheme of Cover and El Gamal [5]. In this coding scheme, the relay compresses its received output Y_1^n as in Wyner–Ziv source coding with the ultimate receiver output Y^n as side information.

In both coding schemes, every bit of relay information carries one bit of information about the channel input, although the relay does not know the channel input. And the relay information can be summarized in a manner completely independent of geometry (random binning) or completely dependent on geometry (random covering).

More surprisingly, we can partition the relay space using both random binning and random covering. Thus, a combination of "hash-and-forward" and "compress-and-forward" achieves the capacity.

The next section proves the achievability using the "hash-and-forward" coding scheme. The "compress-and-forward" scheme is deferred to Section 5 and the combination will be discussed in Sections 6 and 7.

3 Proof of Achievability (Hash and Forward)

We combine the usual random codebook generation with list decoding and random binning of the relay output sequences:

Codebook generation. Generate 2^{nR} independent codewords $X^n(w)$ of length n according to $\prod_{i=1}^n p(x_i)$. Independently, assign all possible relay output sequences in $|\mathcal{Y}_1|^n$ into 2^{nR_0} bins uniformly at random.

Encoding. To send the message index $w \in [2^{nR}]$, the transmitter sends the codeword $X^n(w)$. Upon receiving the output sequence Y_1^n, the relay sends the bin index $b(Y_1^n)$ to the receiver.

Decoding. Let $A_\epsilon^{(n)}$ [6, Section 7.6] denote the set of jointly typical sequences $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ under the distribution $p(x,y)$. The receiver constructs a list

$$L(Y^n) = \{X^n(w) : w \in [2^{nR}], (X^n(w), Y^n) \in A_\epsilon^{(n)}\}$$

of codewords $X^n(w)$ that are jointly typical with Y^n. Since the relay output Y_1 is a deterministic function of (X, Y), then for each codeword $X^n(w)$ in $L(Y^n)$, we can determine the corresponding relay output $Y_1^n(w) = f(X^n(w), Y^n)$ exactly. The receiver declares $\hat{w} = w$ was sent if there exists a unique codeword $X^n(w)$ with the corresponding relay bin index $b(f(X^n(w), Y^n))$ matching the true bin index $b(Y_1^n)$ received from the relay.

Analysis of the probability of error. Without loss of generality, assume $W = 1$ was sent. The sources of error are as follows (see Figure 4):

(a) The pair $(X^n(1), Y^n)$ is not typical. The probability of this event vanishes as n tends to infinity.

(b) The pair $(X^n(1), Y^n)$ is typical, but there is more than one relay output sequence $Y_1^n(w) = f(X^n(w), Y^n)$ with the observed bin index, i.e., $b(Y_1^n(1)) = b(Y_1^n(w))$. By Markov's inequality, the probability of this event is upper bounded by the expected number of codewords in $L(Y^n)$ with the corresponding relay bin index equal to the true bin index $b(Y_1^n(1))$. Since the bin index is assigned independently and uniformly, this is bounded by

$$2^{nR}\, 2^{-n(I(X;Y)-\epsilon)}\, 2^{-nR_0},$$

which vanishes asymptotically as $n \to \infty$ if $R < I(X;Y) + R_0 - \epsilon$.

(c) The pair $(X^n(1), Y^n)$ is typical and there is exactly one $Y_1^n(w)$ matching the true relay bin index, but there is more than one codeword $X^n(w)$ that is jointly typical with Y^n and corresponds to the same relay output Y_1^n, i.e., $f(X^n(1), Y^n) = f(X^n(w), Y^n)$. The probability of this kind of error is upper bounded by

$$2^{nR}\, 2^{-n(I(X;Y,Y_1)-\epsilon)},$$

which vanishes asymptotically if $R < I(X; Y, Y_1) - \epsilon$.

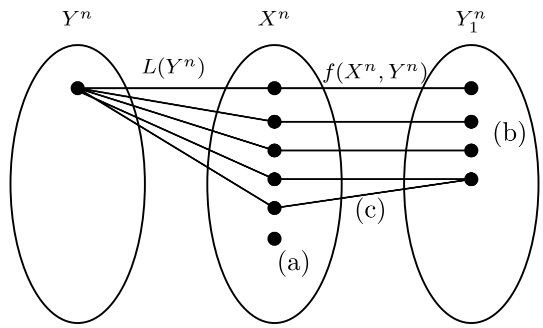

Figure 4: Schematic diagram of "hash-and-forward" coding scheme. The error happens when (a) the true codeword is not jointly typical with Y^n, (b) there is more than one Y_1^n for the same bin index, or (c) there is more than one X^n jointly typical with (Y^n, Y_1^n).

4 Related Work

The general relay channel was introduced by van der Meulen [10]. We refer the readers to Cover and El Gamal [5] for the history and the definition of the general relay channel. For recent progress, refer to Kramer et al. [9], El Gamal et al. [8], and the references therein.

We recall the following achievable rate for the general relay channel investigated in [5].

Theorem 2 ([5, Theorem 7]). *For any relay channel* $(\mathcal{X} \times \mathcal{X}_1, p(y, y_1|x, x, x_1), \mathcal{Y} \times \mathcal{Y}_1)$, *the capacity* C *is lower bounded by*

$$C \geq \sup \min\{I(X; Y, \hat{Y}_1|X_1, U) + I(U; Y_1|X_1, V),\ I(X, X_1; Y) - I(\hat{Y}_1; Y_1|X, X_1, Y, U)\}$$

where the supremum is taken over all joint probability distributions of the form

$$p(u, v, x, x_1, y, y_1, \hat{y}_1) = p(v)p(u|v)p(x|u)p(x_2|v)p(y, y_1|x, x_1)p(\hat{y}_1|x_1, y_1, u)$$

subject to the constraint

$$I(Y_1; \hat{Y}_1|X_1, Y, U) \leq I(X_1; Y|V).$$

Roughly speaking, the achievability of the rate in Theorem 2 is based on a superposition of "decode-and-forward" (in which the relay decodes the message and sends it to the receiver) and "compress-and-forward" (in which the relay compresses its own received signal without decoding and sends it to the receiver). This coding scheme turns out to be optimal for many special cases; Theorem 2 reduces to the capacity when the relay channel is degraded or reversely degraded [5] and when there is feedback from the receiver to the relay [5].

Furthermore, for the semideterministic relay channel with the sender X, the relay sender X_1, the relay receiver $Y_1 = f(X, X_1)$, and the receiver Y, El Gamal and Aref [7] showed that Theorem 2 reduces to the capacity given by

$$C = \max_{p(x,x_1)} \min\{I(X, X_1; Y),\ H(Y_1|X_1) + I(X; Y|X_1, Y_1)\}. \tag{3}$$

Although this setup looks similar to ours, we note that neither (3) nor Theorem 1 implies the other. In a sense, our model is more deterministic in the relay-to-receiver link, while the El Gamal–Aref model is more deterministic in the transmitter-to-relay link.

A natural question arises whether our Theorem 1 follows from Theorem 2 as a special case. We first note that in the coding scheme described in Section 2, the relay does neither "decode" nor "compress", but instead "hashes" its received output. Indeed, as a coding scheme, this "hash-and-forward" appears to be a novel method of summarizing the relay's information. However, "hash-and-forward" is not the unique coding scheme achieving the capacity

$$C(R_0) = \max_{p(x)} \min\{I(X; Y) + R_0,\ I(X; Y_1, Y)\}.$$

In the next section, we show that "compress-and-forward" can achieve the same rate.

5 Compress and Forward

Theorem 1 was proved using "hash-and-forward" in Section 3. Here we argue that the capacity in Theorem 1 can also be achieved by "compress-and-forward".

We start with a special case of Theorem 2. The "compress-and-forward" part (cf. [5, Theorem 6]), combined with the relay-to-receiver communication of rate R_0, gives the achievable rate

$$R^*(R_0) = \sup I(X; Y, \hat{Y}_1), \tag{4}$$

where the supremum is over all joint distributions of the form $p(x)p(y, y_1|x)p(\hat{y}_1|y_1)$ satisfying

$$I(Y_1; \hat{Y}_1|Y) \le R_0. \tag{5}$$

Here the inequality (5) comes from the Wyner–Ziv compression [11] of the relay's output Y_1^n based on the side information Y^n. The achievable rate (4) captures the idea of decoding X^n based on the receiver's output Y^n and the compressed version $\hat{Y}_1^n$ of the relay's output Y_1^n.

We now derive the achievability of the capacity

$$C(R_0) = \max_{p(x)} \min\{I(X;Y,Y_1), I(X;Y) + R_0\}$$

from an algebraic reduction of the achievable rate given by (4) and (5). First observe that, because of the deterministic relationship $Y_1 = f(X,Y)$, we have

$$I(X;\hat{Y}_1|Y) \geq I(Y_1;\hat{Y}_1|Y).$$

Also note that, for any triple (X,Y,Y_1), if $H(Y_1|Y) > R_0$, there exists a distribution $p(\hat{y}_1|y_1)$ such that $(X,Y) \to Y_1 \to \hat{Y}_1$ and $I(Y_1;\hat{Y}_1|Y) = R_0$.

Henceforth, maximums are taken over joint distributions of the form $p(x)p(y,y_1|x)p(\hat{y}_1|y_1)$ with $Y_1 = f(X,Y)$. We have

$$\begin{aligned}
R^*(R_0) &= \sup\{I(X;Y,\hat{Y}_1) : I(Y_1;\hat{Y}_1|Y) \leq R_0\} \\
&\geq \sup\{I(X;Y,\hat{Y}_1) : I(Y_1;\hat{Y}_1|Y) \leq R_0, H(Y_1|Y) > R_0\} \\
&\geq \sup\{I(X;Y,\hat{Y}_1) : I(Y_1;\hat{Y}_1|Y) = R_0, H(Y_1|Y) > R_0\} \\
&= \sup\{I(X;Y) + I(X;\hat{Y}_1|Y) : I(Y_1;\hat{Y}_1|Y) = R_0, H(Y_1|Y) > R_0\} \\
&\geq \sup\{I(X;Y) + I(Y_1;\hat{Y}_1|Y) : I(Y_1;\hat{Y}_1|Y) = R_0, H(Y_1|Y) > R_0\} \\
&= \sup\{I(X;Y) + R_0 : I(Y_1;\hat{Y}_1|Y) = R_0, H(Y_1|Y) > R_0\} \\
&= \max\{I(X;Y) + R_0 : H(Y_1|Y) > R_0\}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
R^*(R_0) &= \sup\{I(X;Y,\hat{Y}_1) : I(Y_1;\hat{Y}_1|Y) \leq R_0\} \\
&\geq \sup\{I(X;Y,\hat{Y}_1) : I(Y_1;\hat{Y}_1|Y) \leq R_0, H(Y_1|Y) \leq R_0\} \\
&\geq \sup\{I(X;Y,\hat{Y}_1) : \hat{Y}_1 = Y_1, H(Y_1|Y) \leq R_0\} \\
&= \sup\{I(X;Y,Y_1) : \hat{Y}_1 = Y_1, H(Y_1|Y) \leq R_0\} \\
&= \max\{I(X;Y,Y_1) : H(Y_1|Y) \leq R_0\}.
\end{aligned}$$

Thus, we have

$$R^*(R_0) \geq \max_{p(x):H(Y_1|Y)>R_0} I(X;Y) + R_0$$

and

$$R^*(R_0) \geq \max_{p(x): H(Y_1|Y) \leq R_0} I(X; Y, Y_1),$$

and therefore,

$$R^*(R_0) \geq \max_{p(x)} \min\{I(X;Y) + R_0,\ I(X;Y,Y_1)\}.$$

In words, "compress-and-forward" achieves the capacity.

6 Discussion: Random Binning vs. Random Covering

It is rather surprising that both "hash-and-forward" and "compress-and-forward" optimally convey the relay information to the receiver, especially because of the dual nature of compression (random covering) and hashing (random binning). (And the hashing in "hash-and-forward" should be distinguished from the hashing in Wyner–Ziv source coding.) The example in Figure 5 illuminates the difference between the two coding schemes.

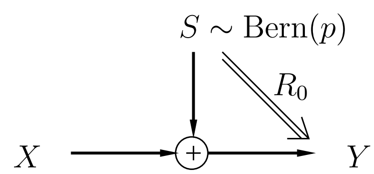

Figure 5: Binary symmetric channel with rate-limited state information at receiver.

Here the binary input $X \in \{0, 1\}$ is sent over a binary symmetric channel with cross-over probability p, or equivalently, the channel output $Y \in \{0, 1\}$ is given as

$$Y = X + S \pmod 2,$$

where the binary additive noise $S \sim \text{Bern}(p)$ is independent of the input X. With no information on S available at the transmitter or the receiver, the capacity is

$$C(0) = 1 - H(p).$$

Now suppose there is an intermediate node which observes S and "relays" that information to the decoder through a side channel of rate R_0. Since $S = X + Y$ is a deterministic function

of (X, Y), Theorem 1 applies and we have

$$C(R_0) = 1 - H(p) + R_0$$

for $0 \leq R_0 \leq H(p)$.

There are two ways of achieving the capacity. First, hashing. The relay hashes the entire binary $\{0,1\}^n$ into 2^{nR_0} bins, then sends the bin index $b(S^n)$ of S^n to the decoder. The decoder checks whether a specific codeword $X^n(w)$ is typical with the received output Y^n and then whether $S^n(w) = X^n(w) + Y^n$ matches the bin index.

Next, covering. The relay compresses the state sequence S^n using the binary lossy source code with rate R_0. More specifically, we use the standard backward channel for the binary rate distortion problem (see Figure 6):

$$S = \hat{S} + U.$$

Here $\hat{S} \in \{0,1\}$ is the reconstruction symbol and $U \sim \text{Bern}(q)$ is independent of $\hat{S}$ (and X)

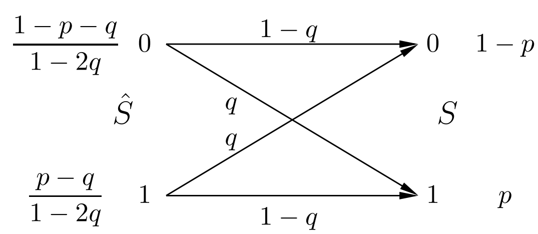

Figure 6: Backward channel for the binary rate distortion problem.

with parameter q satisfying

$$R_0 = I(S; \hat{S}) = H(p) - H(q).$$

Thus, using nR_0 bits, the ultimate receiver can reconstruct $\hat{S}^n$.

Finally, decoding $X^n \sim \text{Bern}(1/2)$ based on $(Y^n, \hat{S}^n)$, we can achieve the rate

$$\begin{aligned} I(X; Y, \hat{S}) &= I(X; X+S, S+U) \\ &\geq I(X; X+U) \\ &= 1 - H(q) \\ &= 1 - H(p) + R_0. \end{aligned}$$

In summary, the optimal relay can partition its received signal space into either random bins or Hamming spheres.

The situation is somewhat reminiscent of that of lossless block source coding. Suppose $\{X_i\}$ is independent and identically distributed (i.i.d.) $\sim$ Bern(p). Here are two basic methods of compressing X^n into $nH(p)$ bits with asymptotically negligible error.

1) *Hashing.* The encoder simply hashes X^n into one of $2^{nH(p)}$ indices. With high probability, there is a unique typical sequence with matching hash index.

2) *Enumeration* [3]. The encoder enumerates $2^{nH(p)}$ typical sequences. Then $nH(p)$ bits are required to give the enumeration index of the observed typical sequence. With high probability, the given sequence X^n is typical.

While these two schemes are apparently unrelated, they are both extreme cases of the following coding scheme.

3) *Covering with hashing.* By fixing $p(\hat{x}|x)$ and generating independent sequences $\hat{X}^n(i)$, $i = 1, \ldots, 2^{nI(X;\hat{X})}$, each i.i.d. $\sim p(\hat{x})$, we can induce a set of $2^{nI(X;\hat{X})}$ coverings for the space of typical X^n's. For each cover $\hat{X}^n(i)$, there are $\approx 2^{nH(X|\hat{X})}$ sequences that are jointly typical with $\hat{X}^n(i)$. Therefore, by hashing X^n into one of $2^{nH(X|\hat{X})}$ hash indices and sending it along the cover index, we can recover a typical X^n with high probability. This scheme requires $n(I(X;\hat{X}) + H(X|\hat{X})) = nH(p)$ bits.

Now if we take $\hat{X}$ independent of X, then we have the case of hashing only. On the other hand, if we take $\hat{X} = X$, then we have enumeration only, in which case the covers are Hamming spheres of radius zero. It is interesting to note that the combination scheme works under any $p(\hat{x}|x)$.

Thus motivated, we combine "hash-and-forward" with "compress-and-forward" in the next section.

7 Compress, Hash, and Forward

Here we show that a combination of "compress-and-forward" and "hash-and-forward" can achieve the capacity

$$C(R_0) = \max_{p(x)} \min\{I(X;Y,Y_1),\ \ I(X;Y) + R_0\}$$

for the setup in Theorem 1.

We first fix an *arbitrary* conditional distribution $p(\hat{y}_1|y_1)$ and generate $2^{n(I(Y_1;\hat{Y}_1)+\epsilon)}$ sequences $\hat{Y}_1^n(i)$, $i = 1, 2, \ldots, 2^{n(I(Y_1;\hat{Y}_1)+\epsilon)}$, each i.i.d. $\sim p(\hat{y}_1)$. Then, with high probability, a typical Y_1^n has a jointly typical cover $\hat{Y}_1^n(Y_1^n)$. (If there is more than one, pick the one with the smallest index. If there is none, assign $\hat{Y}_1^n(1)$.)

There are two cases to consider, depending on our choice of $p(\hat{y}_1|y_1)$ (and the input codebook distribution $p(x)$). First suppose

$$I(Y_1; \hat{Y}_1|Y) \geq R_0. \tag{6}$$

If we treat $\hat{Y}_1^n(Y_1^n)$ as the relay output, $\hat{Y}_1^n$ is a deterministic function of Y_1^n and thus of (X^n, Y^n). Therefore, we can use "hash-and-forward" on $\hat{Y}_1^n$ sequences. (Markov lemma [2] justifies treating $\hat{Y}_1^n(Y_1^n)$ as the output of the memoryless channel $p(y, \hat{y}_1|x)$.) This implies that we can achieve

$$R^*(R_0) = \min\{I(X;Y) + R_0,\ \ I(X;Y,\hat{Y}_1)\}.$$

But from (6) and the functional relationship between Y_1 and (X, Y), we have

$$\begin{aligned} I(X;Y,\hat{Y}_1) &= I(X;Y) + I(X;\hat{Y}_1|Y) \\ &\geq I(X;Y) + I(Y_1;\hat{Y}_1|Y) \\ &\geq I(X;Y) + R_0. \end{aligned}$$

Therefore,

$$R^*(R_0) = I(X;Y) + R_0,$$

which is achieved by the above "compress-hash-and-forward" scheme with $p(x)$ and $p(\hat{y}_1|y_1)$ satisfying (6).

Alternatively, suppose

$$I(Y_1; \hat{Y}_1|Y) \leq R_0. \tag{7}$$

Then, we can easily achieve the rate $I(X;Y,\hat{Y}_1)$ by the "compress-and-forward" scheme. The rate $R_0 \geq I(Y_1;\hat{Y}_1|Y)$ suffices to convey $\hat{Y}_1^n$ to the ultimate receiver.

But we can do better by using the remaining $\Delta = R_0 - I(Y_1;\hat{Y}_1|Y)$ bits to further hash Y_1^n itself. (This hashing of Y_1^n should be distinguished from that of Wyner–Ziv coding which bins $\hat{Y}_1^n$ codewords.) By treating $(Y, \hat{Y}_1)$ as a new ultimate receiver output and Y_1 as the relay output, "hash-and-forward" on top of "compress-and-forward" can achieve

$$R^*(R_0) = \min\{I(X;Y,\hat{Y}_1) + \Delta,\ \ I(X;Y,\hat{Y}_1,Y_1)\}. \tag{8}$$

Since

$$\begin{aligned} I(X;Y,\hat{Y}_1) + \Delta &= I(X;Y,\hat{Y}_1) - I(Y_1;\hat{Y}_1|Y) + R_0 \\ &\geq I(X;Y,\hat{Y}_1) - I(X;\hat{Y}_1|Y) + R_0 \\ &= I(X;Y) + R_0 \end{aligned}$$

and

$$I(X;Y,\hat{Y}_1,Y_1) = I(X;Y,Y_1),$$

the achievable rate in (8) reduces to

$$R^*(R_0) = \min\{I(X;Y) + R_0,\ \ I(X;Y,Y_1)\}.$$

Thus, by maximizing over input distributions $p(x)$, we can achieve the capacity for either case (6) or (7).

It should be stressed that our combined "compress-hash-and-forward" is optimal, regardless of the covering distribution $p(\hat{y}_1|y_1)$. In other words, any covering (geometric partitioning) of Y_1^n space achieves the capacity if properly combined with hashing (nongeometric partitioning) of the same space. In particular, taking $\hat{Y}_1 = Y_1$ leads to "hash-and-forward" while taking the optimal covering distribution $p^*(\hat{y}_1|y_1)$ for (4) and (5) in Section 5 leads to "compress-and-forward".

8 Ahlswede–Han Conjecture

In this section, we show that Theorem 1 confirms the following conjecture by Ahlswede and Han [1] on the capacity of channels with rate-limited state information at the receiver, for the special case in which the state is a deterministic function of the channel input and the output.

First, we discuss the general setup considered by Ahlswede and Han, as shown in Figure 7. Here we assume that the channel $p(y|x,s)$ has independent and identically distributed state

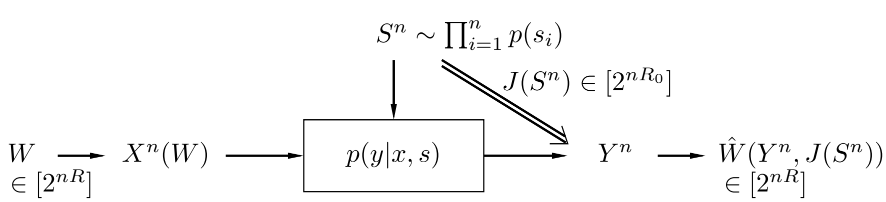

Figure 7: Channel with rate-limited state information at the decoder.

S^n and the decoder can be informed about the outcome of S^n via a separate communication channel at a fixed rate R_0. Ahlswede and Han offered the following conjecture on the capacity of this channel.

Conjecture (Ahlswede–Han [1, Section V]). *The capacity of the state-dependent channel $p(y|x,s)$ as depicted in Figure 7 with rate-limited state information available at the receiver via a separate communication link of rate R_0 is given by*

$$C(R_0) = \max I(X;Y|\hat{S}), \tag{9}$$

where the maximum is over all joint distributions of the form $p(x)p(s)p(y|x,s)p(\hat{s}|s)$ such that

$$I(S;\hat{S}|Y) \le R_0$$

and the auxiliary random variable $\hat{S}$ has cardinality $|\hat{\mathcal{S}}| \le |\mathcal{S}| + 1$.

It is immediately seen that this problem is a special case of a relay channel with a noiseless link (Figure 3). Indeed, we can identify the relay output Y_1 with the channel state S and identify the relay channel $p(y, y_1|x) = p(y_1|x)p(y|x, y_1)$ with the state-dependent channel $p(s)p(y|x,s)$. Thus, the channel with rate-limited state information at the receiver is a relay channel in which the relay channel output Y_1 is independent of the input X. The binary symmetric channel example in Section 6 corresponds to this setup.

Now when the channel state S is a deterministic function of (X, Y), for example, $S = X+Y$ as in the binary example in Section 6, Theorem 1 proves the following.

Theorem 3. *For the state-dependent channel $p(y|x,s)$ with state information available at the decoder via a separate communication link of rate R_0, if the state S is a deterministic function of the channel input X and the channel output Y, then the capacity is given by*

$$C(R_0) = \max_{p(x)} \min\{I(X;Y) + R_0,\ I(X;Y,S)\}. \tag{10}$$

Our analysis of "compress-and-forward" coding scheme in Section 5 shows that (9) reduces to (10), confirming the Ahlswede–Han conjecture when S is a function of (X, Y). On the other hand, our proof of achievability (Section 3) shows that "hash-and-forward" is equally efficient for informing the decoder of the state information.

9 Concluding Remarks

Even a completely oblivious relay can boost the capacity to the cut set bound, if the relay reception is fully recoverable from the channel input and the ultimate receiver output. And

there are two basic alternatives for the optimal relay function—one can either compress the relay information as in the traditional method of "compress-and-forward," or simply hash the relay information. In fact, infinitely many relaying schemes that combine hashing and compression can achieve the capacity. While this development depends heavily on the deterministic nature of the channel, it reveals an interesting role of hashing in communication.

References

[1] R. Ahlswede and T. S. Han, "On source coding with side information via a multiple-access channel and related problems in multi-user information theory," *IEEE Trans. Inform. Theory*, vol. IT-29, no. 3, pp. 396–412, 1983.

[2] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications*, G. Longo, Ed. New York: Springer-Verlag, 1978.

[3] T. M. Cover, "Enumerative source encoding," *IEEE Trans. Inform. Theory*, vol. IT-19, no. 1, pp. 73–77, Jan. 1973.

[4] ——, "The capacity of the relay channel," in *Open Problems in Communication and Computation*, T. M. Cover and B. Gopinath, Eds. New York: Springer-Verlag, 1987, pp. 72–73.

[5] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. IT-25, no. 5, pp. 572–584, Sept. 1979.

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.

[7] A. El Gamal and M. Aref, "The capacity of the semideterministic relay channel," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 3, p. 536, May 1982.

[8] A. El Gamal, N. Hassanpour, and J. Mammen, "Relay networks with delay," submitted to *IEEE Trans. Inform. Theory,* 2006.

[9] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inform. Theory*, vol. IT-51, no. 9, pp. 3037–3063, Sept. 2005.

[10] E. C. van der Meulen, "Three-terminal communication channels," *Adv. Appl. Prob.*, vol. 3, pp. 120–154, 1971.

[11] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-22, no. 1, pp. 1–10, 1976.

[12] Z. Zhang, "Partial converse for a relay channel," *IEEE Trans. Inform. Theory*, vol. IT-34, no. 5, pp. 1106–1110, Sept. 1988.